面向网络游记时间特征的情感分析模型*

程翠琼 徐 健
(中山大学资讯管理学院 广州 510006)

摘要: 【目的】通过对网络游记进行情感分析, 发现游客对旅游地情感倾向的时间分布规律。【应用背景】越来越多人通过浏览大量网络游记来收集信息, 制定旅游计划。网络游记成为旅游者搭配旅游地及出游时间的重要参考内容, 也为商家提供了商机。【方法】提出面向网络游记时间特征的情感分析模型, 分析游客情感的时间变化规律。该模型包括 5 个模块: 网络游记文本内容及旅游时间数据的采集、游记文本预处理、情感标注、按时间段统计游记情感特征分值、游记情感时间特征分析。并从网络抓取 4 种类型旅游地游记对模型进行实验。【结果】在 7 类情感中, [好]的情感均值在各旅游地的各月份中总是远高于其他情感, 较为稳定; [好]、[乐]和[恶]在不同月份的波动程度较大; 情感随时间的波动与相应游记数量并不相关, 即传统的旅游地旺季和淡季的划分与游客的实际情感体验并不相关。【结论】该模型能够有效地反映旅游地的游客情感随时间变化的波动, 进而为旅游管理者、潜在旅游者信息获取提供新的信息参考渠道。

关键词: 网络游记 情感分析 情感词典 时间特征
分类号: G350

1 引 言

由于旅游地景点的季节波动以及旅游旺淡季人流量或服务质量不平衡都会影响到游客的情感, 因此不同时间段的旅游体验不一样, 游客对旅游地的情感倾向也会有所变化。随着社会媒体的发展, 越来越多旅游者通过网络游记分享旅游体验, 表达自己对旅游地的意见及情感倾向。网络游记数据提供了游记文本及旅游出发时间信息, 能准确反映出在不同时间段游客对旅游地的意见及情感倾向。而针对网络游记进行基于时间特征的分析, 结合细致的情感分类, 可发现游客各类情感随时间变化的规律, 选取关键性情感指标, 一方面能为旅游地管理者提供新的信息源, 发现积极情感强烈但游客数量不多的时间段, 从而调整宣传策略;而为了让游客在网络上传播旅游地积极的口碑, 旅游地管理者可发现消极情感强烈的时间段, 进而优化该时间段的旅游资源分配。另一方面, 可以为潜在旅游者提供根据不同时间段的游客情感倾向的旅游地检索新渠道, 满足更多元的检索需求, 让潜在旅游者可以根据自身的旅游期望选择最佳的旅游地和旅游时间的组合。

本文利用基于情感词典的情感分析方法, 对游客发表的网络游记进行情感标注, 通过分析不同时间段游记的情感波动情况, 发现游客对旅游地情感倾向随时间变化的规律, 进而为旅游管理者、潜在旅游者信息获取提供新的信息参考渠道。

2 相关研究

情感分析又称意见挖掘, 是指通过对用户发表的内容文本进行主客观性观点、情绪、极性的分析和挖掘, 判断出文本的情感倾向分类[1]。情感分析的方法主要分为两大类: 基于情感词典的方法和基于机器学习的方法。

(1) 基于情感词典进行情感分析, 要构建一个能

通讯作者: 徐健, ORCID: 0000-0003-4886-4708, E-mail: issxj@mail.sysu.edu.cn。

*本文系广东省科技专项项目“基于内容的科技文献分析服务平台”(项目编号: 2016B030303003)和广东省科技专项项目“面向主题的中文语料库构建方法与技术”(项目编号: 22015A030401037)的研究成果之一。

准确判断情感词所属的类别及情感值的情感词典，情感词典的质量对情感分析的效果影响很大。Hu 等[2]通过 WordNet 得到情感词及其情感极性，根据句子正负情感分的离差决定该句子的情感倾向。在中文领域中，朱嫣岚等[3]使用 HowNet 进行语义相似度和语义相关场的计算，提出两种词汇情感倾向性计算方法，判别准确率高。史伟等[4]构建了模糊情感词本体，进行微博的情感计算，得到公众对突发事件的情感曲线。

(2) 基于机器学习的情感分析可归结为文本分类问题，常用的机器学习模型有：朴素贝叶斯(Naive Bayes)，最大熵(Maximum Entropy)，中心向量分类法，K 最近邻分类和支持向量机(SVM)。Pang 等[5]分别使用朴素贝叶斯、最大熵、支持向量机三种方法对电影评论进行情感倾向性分类，发现支持向量机方法的分类效果最佳，精准度可达 80%。在中文领域中，徐军等[6]以新闻及评论为语料，采用朴素贝叶斯和最大熵方法进行情感分类，发现在多数情况下最大熵方法的准确率高于贝叶斯方法。

情感分析在旅游领域的应用主要在旅游服务评论以及景点评论上。在旅游服务评论方面，Banić 等[7]利用 KNIME 对酒店的在线评论进行情感分析，得到酒店某一特征的情感分值及酒店整体情感倾向，为潜在客户提供决策参考，也为酒店管理人员指明改进方向。在旅游目的地景点评论方面，王素格等[8]利用词对间的依存关系，设计具有情感倾向的“特征–观点”对抽取算法，并对山西旅游景点评论进行“特征–观点”对的抽取，有效挖掘游客的观点信息。郑文英[9]面向旅行目的地在线评论信息，基于逐点语义分析法提取情感倾向词组，从三大方面分析游客反映的关于旅游地的问题，并提出建议。

游客对旅游地的情感不仅因人而异，也因时而异。金程[10]通过半结构化日记发现游客在不同旅游阶段的情感强度呈动态变化。胡传东[11]利用 ROST CM6 对骑行游记进行情感分析，并根据情感倾向的起伏变化规律，找出骑行体验中高峰体验的路段。于静[12]构建基于微博大数据的游客情感分析模型，对西安市区内的用户微博进行情感分析，分析了游客情感的时间变化特征。Li 等[13]研究并设计了一个用于旅游网络观点的视觉分析系统 VisTravel，通过对网络游记及评论的情感分析，能有效展示游客情感随时间的变化。

情感分析在旅游领域仍有较大研究和应用空间。目前研究多数仅基于文本情感二元划分方法，分为积极和消极情感，且多以分析旅游服务评论为主。与之相比，本研究将时间特征引入到游客对旅游目的地情感倾向的研究中，使用更为细致的文本情感分类，构建面向网络游记时间特征的情感分析模型，因此能够更好地反映游客对旅游目的地情感的时间变化规律。

3　面向网络游记时间特征的情感分析模型

通过相关研究发现，当前在旅游领域的情感分析主要基于旅游产品或旅游服务的用户评论，本文认为可以对网络游记进行深入挖掘，引入时间特征，发现游客情感随时间的变化特征，从而能结合时间维度为潜在旅游者提供旅游决策参考，同时有助于为旅游地提供全年不同时间段的旅游产品策划、营销策略等建议。本文提出的面向网络游记时间特征的情感分析模型是基于网络游记的文本内容，对网络游记文本进行情感分析，研究游客对旅游目的地情感的时间特征变化，该情感分析模型如图 1 所示。

面向网络游记时间特征的情感分析处理过程，可以分为 5 大模块：采集网络游记数据、对游记文本进行预处理、情感标注、计算游记的情感特征分值、游记的情感时间特征结果分析。

(1) 采集网络游记文本内容及旅游时间数据。先设定实验数据的时间范围以及旅游目的地，再利用网络爬虫采集符合以下特征的游记数据：数据要素齐全(旅游目的地、标题、出游时间、文本内容)；游记显示的出游时间在选定的时间段内。

(2) 游记文本预处理。采集的部分游记会存在描写多个旅游地的情况，可能导致目标旅游地的情感倾向性不同。因此，为了提高情感的目标指向性，降低同游的旅游地对目标旅游地情感的影响，要对游记数据进行预处理，剔除包含不同旅游地的游记数据。再进行信息提取，主要提取游记标题、出游时间、文本内容等作为信息元，存储到数据库中，作为实验数据集。

(3) 情感标注。构建适用的情感词典，将情感词典中的情感词逐一与实验数据集中的游记文本内容匹配，记录下游记中出现该情感词的位置。由于文章中会存在否定句，对句子或文章进行情感类别识别时，

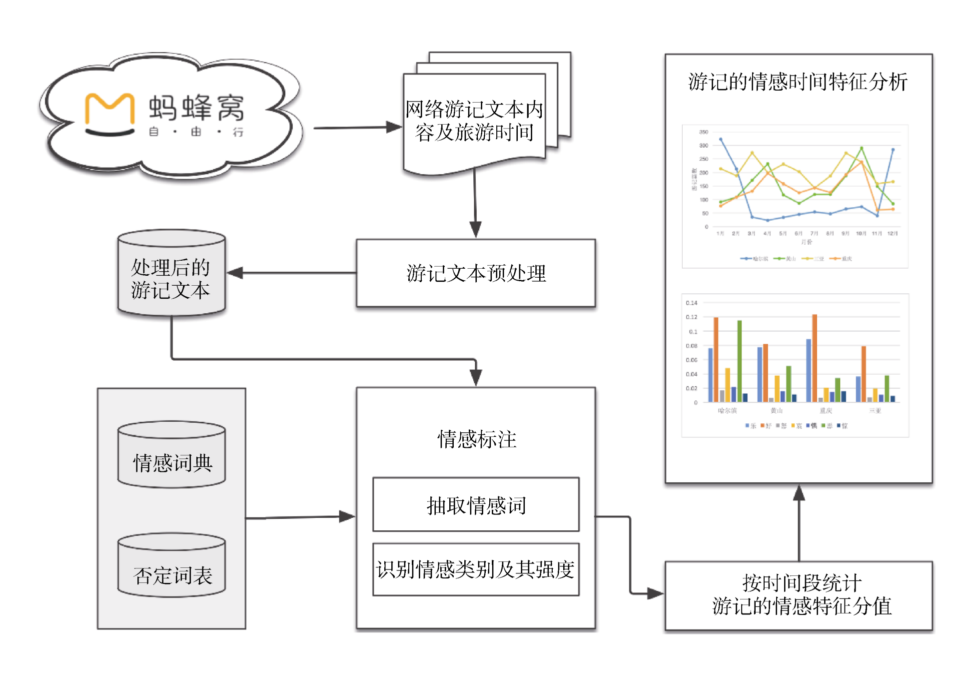


图 1　面向网络游记时间特征的情感分析模型

还需考虑是否存在影响情感类别的否定词。因此，先识别情感词在情感词典中所属类别以及其对应的情感强度，再判断抽取的情感词前是否存在否定词，若存在否定词，则进行相应的情感迁移，最后记录下识别到的情感类别及其强度。

(4) 按时间段统计游记的情感特征分值。对游记文本中识别到的所有情感类型及其强度进行汇总，分别计算该游记文本中各种情感类别的强度之和，从而计算出该篇游记文本中各种情感类别的平均分值。再选择一个合适的时间单位(季度或月份或星期等)，对全年中各个时间段的所有游记文本中各种情感类别的平均分值进行叠加，得到各个时间段整体游记的不同情感类别的情感分值，从而计算出各个时间段内整体游记的不同情感类别的情感平均分值。

(5) 游记的情感时间特征分析。对得到的不同旅游地的情感数据进行分析，得到不同时间段游记的情感波动情况，发现游客对旅游地情感倾向随时间变化的规律，进而为旅游管理者、潜在旅游者信息获取提供新的信息参考渠道。

4　面向网络游记时间特征的情感分析实验

为验证模型的有效性，本文从蚂蜂窝网[12]采集相关的游记数据，通过对典型的不同类型旅游地的分析，以期发现基于时间特征的各个旅游地的情感规律。

4.1　网络游记数据采集及预处理

为使研究结果更具有代表性，本研究选择 4 个分布在不同区域且各具特色的旅游城市，分别是东北资源型的哈尔滨、自然景观型的黄山、都市观光型的重庆以及南方海滨型的三亚。选用爬虫工具“火车采集器”[14]抓取了蚂蜂窝[15]上的网络游记数据作为情感分析的实验数据集，数据采集时间为: 2016 年 11 月。实验数据时间范围设定在 2013 年 1 月 1 日至 2015 年 12 月 31 日，采集到哈尔滨游记 1 347 篇、黄山游记 1 671 篇、重庆游记 1 916 篇、三亚游记 2 349 篇。由于采集的游记中存在描写多个旅游地的情况，为了保证情感分析的准确度，对数据中包含不同旅游地的游记进行剔除。最终得到哈尔滨游记 821 篇、黄山游记 969 篇、重庆游记 1 234 篇和三亚游记 2 183 篇。将采集到的网络游记数据中出现的旅游目的地、游记内容及对应的旅游时间进行结构化抽取，存入数据库中。

4.2　情感标注

(1) 选择情感词典

与基于机器学习的方法相比，基于词典的倾向性评分方法则简单且高效，适合工程应用，但此方法的效果受情感词典的质量影响较大[16]。目前，在中文领域，主要选用的情感词典有知网的情感分析用词语集

(HowNet)①，台湾大学中文通用情感词典(NTUSD)②以及中文情感词汇本体库③。

本文选用 2012 年大连理工大学信息检索研究室发布的中文情感词汇本体库作为情感词典。该词典将情感共分为乐、好、怒、哀、惧、恶、惊 7 大类，包含 27 352 个中文情感词汇。

(2) 情感标注过程

将中文情感词汇本体库中的情感词逐一与实验数据集中的游记文本内容进行匹配，得到该情感词在该游记中出现的位置。对于游记中情感词被否定词修饰的情况，采用杜振雷[17]提出的否定词修饰后情感迁移的方法。若情感词前的 5 个字符中存在否定词，则进行相应的情感迁移，最后记录下识别到的情感类别及其强度，再进行下一个情感词的匹配。

例如，"说实话，饺子和面条都不怎么好吃，要不是醋和辣椒，我都吃不下去"中出现情感词"好吃"，根据中文情感词汇本体库查到属于情感大类中的[好]，表示对食物的赞扬，情感强度为 3。由于在"好吃"前出现了否定词"不"，因此进行情感迁移，情感迁移为[恶]，情感强度被减弱，降为 0.6。

4.3 计算情感特征分值

本文面向网络游记文本对每类情感都提取两个情感特征指标，分别是单个月份中网络游记的情感均值和各种情感在月份上的情感标准差。单个月份中网络游记的情感均值反映了以该月份所有的网络游记为整体的情感倾向，各种情感在月份上的情感标准差表现了情感随月份的波动程度。

由于每篇网络游记都包含情感词，因此均为主观文本，可进行情感特征的量化。将网络游记文本表示为篇章集 $D=\{D_1, D_2, \cdots, D_n\}$，按月份分别对 7 类情感进行情感特征分值计算。由于每类情感的计算方法相同，下面以"乐"(Joy)为例介绍情感特征分值的计算方法。

(1) 情感均值能去除网络游记篇数的影响，较准确地反映单个月份中网络游记整体的情感特征。对月份 Mon_i 中所有网络游记[乐]的情感均值进行平均，得到月份 Mon_i [乐]的情感均值，表示为 $MonAvgJoy_i$，如公式(1)所示。

$$MonAvgJoy_i = \frac{\sum_{k=1}^{n} AvgJoy_k}{n} \tag{1}$$

其中，$AvgJoy_k$ 为月份 Mon_i 中的网络游记 D_k [乐]的情感分均值，即网络游记 D_k 中情感类别为[乐]的情感词的情感强度之和与该游记中 7 类情感词次数之和的商。n 为月份 Mon_i 中网络游记的篇数。

(2) 各类情感在各月份的情感均值标准差能反映游客情感在各月份的波动程度。使用标准差公式[18]计算各个月份[乐]的情感标准差，能较准确地反映[乐]的情感在各月份的波动程度，记为 $StdJoy$，如公式(2)所示。

$$StdJoy = \sqrt{\frac{\sum_{i=1}^{12}\left(MonScoJoy_i - \frac{1}{12}\sum_{k=1}^{12} MonScoJoy_k\right)^2}{12}} \tag{2}$$

其中，$MonScoJoy_i$ 为月份 Mon_i [乐]的情感分值。

5 实验结果分析

根据第 4 节的处理方法对实验数据集中的网络游记内容文本进行情感分析，得到哈尔滨、黄山、重庆、三亚这 4 个不同类型旅游地的情感特征数据。在这些数据的基础上，对各旅游地网络游记数量的时间分布情况以及各旅游地网络游记情感时间变化情况进行分析，从中可以发现游客情感的时间特征规律，进而为旅游管理者、潜在旅游者信息获取提供新的信息参考渠道。

5.1 网络游记数量的时间分布

从网络游记文本内容中提取出明确的出游时间，得到各月份游记的数量，在一定程度上能反映出当地的游客数量情况。哈尔滨、黄山、重庆、三亚旅游人数的时间变化情况，如图 2 所示。

①http://www.keenage.com/html/c_index.html.

②http://nlg18.csie.ntu.edu.tw: 8080/opinion/index.html.

③http://ir.dlut.edu.cn/EmotionOntologyDownload.

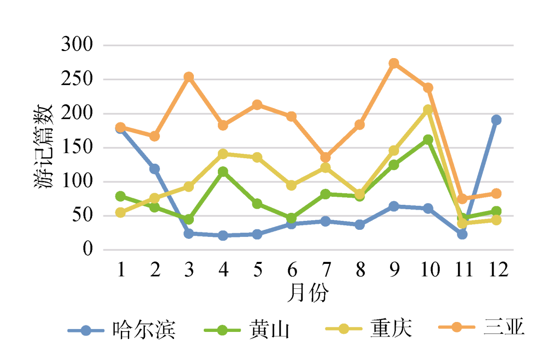


图 2　2013 年–2015 年各城市各月份的网络游记篇数

从整体看各旅游地游客的月度变化，发现 9 月、10 月出游人数相对较多，11 月、12 月较少，与我国法定假期放假情况相对应。不同旅游地最高点的月份不完全相同，随时间变化趋势不一样，说明游客会根据不同月份计划出游的旅游地。整体分布上，各旅游地的网络游记分布情况区别很大。其中，游记数量最多的是 9 月份的三亚，多达 274 篇，可见 9 月的三亚极其受游客欢迎。游记数量最少的是 4 月份的哈尔滨，只有 21 篇，而 12 月份的哈尔滨，高达 191 篇，可见哈尔滨游客数量随时间波动很大，冬季之外的哈尔滨对游客吸引力不足。黄山和重庆的游记数量分布相似，均呈“M”字型，春秋两季为旅游旺季，与当地气候关系密切。相比之下，三亚的游记数量分布都比较高，只有 11 月、12 月相对偏低，说明旅游开发较成熟。

5.2　旅游地网络游记情感的时间变化分析

以月为时间单位对游客情感的时间特征规律进行探索。

(1) 各类情感的情感均值范围

对各旅游目的地游客的网络游记情感变化情况绘制情感变化折线图，如图 3 所示。

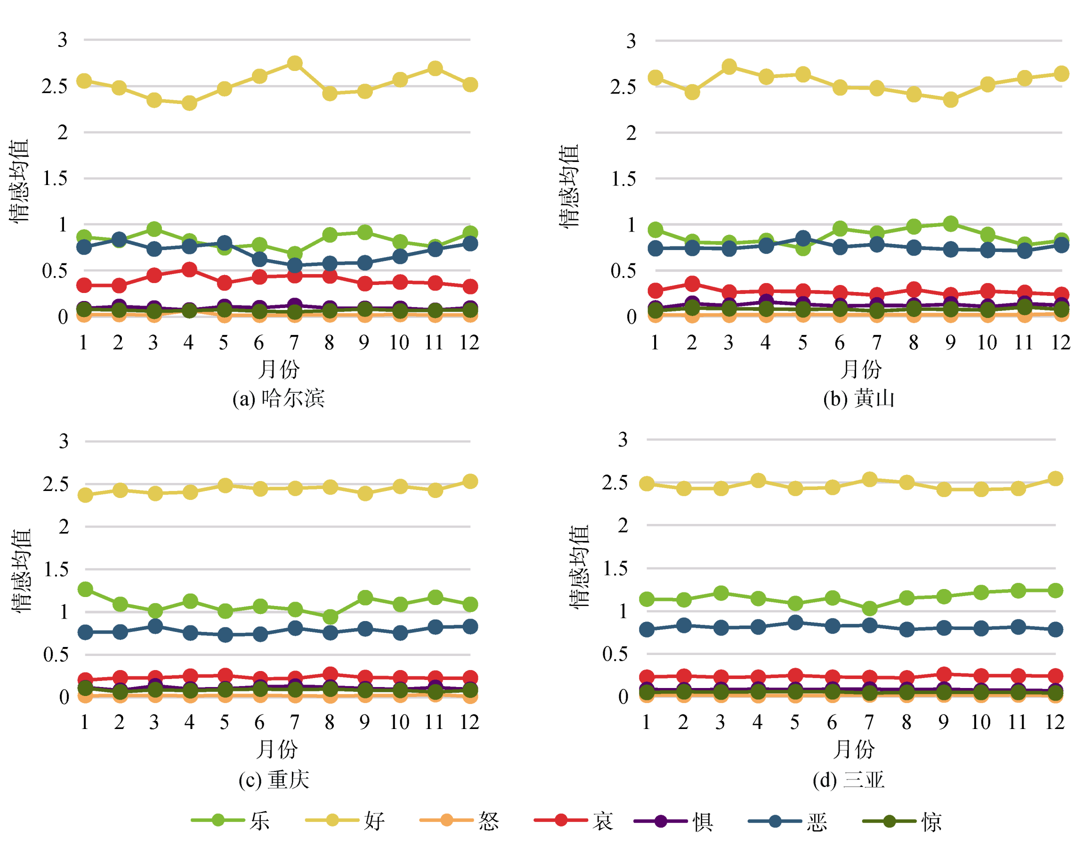


图 3　2013 年–2015 年各城市各月份的网络游记情感均值

从图 3 整体可以看到东北资源型的哈尔滨、自然景观型的黄山、都市观光型的重庆以及南方海滨型的三亚这 4 个不同类型的旅游地情感月度变化情况。[好]的情感均值在各旅游地中都分布较高，其次是[乐]和[恶]，都在 1 上下浮动，其他 4 种情感在 0.5 以下。[好]的情感包括尊敬、赞扬、喜爱、祝愿等，说明游客对旅游地的总体感受倾向于喜爱与赞扬，达到了其对旅游度假的期望，同时反映了游客写网络游记时对旅游

地的回忆多为美好的，对积极情感的描写较多，旅游地能通过游客在网络上传递比较好的口碑。

因此，网络游记情感的时间变化特征一可归纳为：在 7 类情感中，[好]的情感均值在各旅游地的各月份中总是最高，远高于其他情感，且较为稳定。

(2) 情感随时间的波动程度

对各城市网络游记情感均值标准差进行计算，结果如图 4 所示。

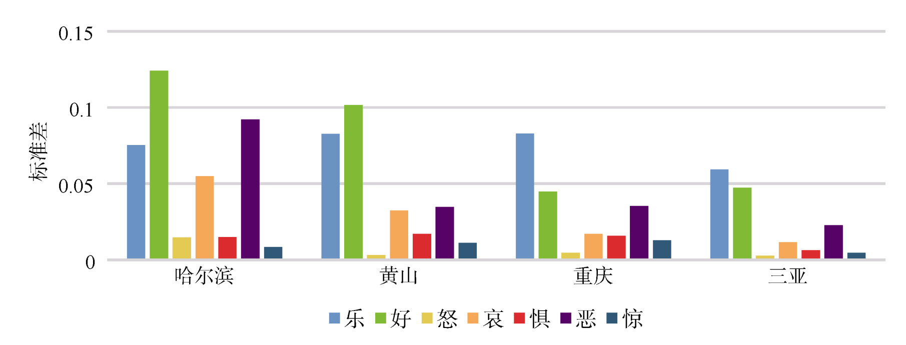


图 4　2013 年–2015 年各城市网络游记情感均值在各月份的标准差

各旅游地的情感标准差分布情况类似，情感[好]、[乐]和[恶]的标准差在各旅游地中都较为明显，而[怒]、[哀]、[惧]、[惊]的标准差相对较小。这说明在分析游客对旅游地情感的时间变化规律时，7 类情感的表达能力相差较大。

由此总结得到网络游记情感的时间变化特征二为：在 7 类情感中，情感[好]、[乐]和[恶]在不同月份的波动程度较大，相比之下[怒]、[哀]、[惧]、[惊]的波动很小。

(3) 情感随时间的波动情况与游记数量变化之间的关系

考虑[好]、[乐]、[恶]这三种明显波动的情感能在一定程度上反映不同月份游客的情感变化情况，因此针对各个旅游地的这三种情感在时间上的变化进行分析。以哈尔滨为例，对[好]、[乐]和[恶]的情感均值及游记数进行分析，如图 5 所示。

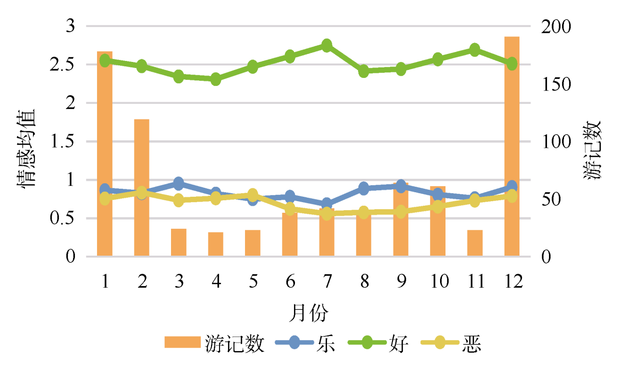


图 5　2013 年–2015 年哈尔滨网络游记情感均值与游记数

游客对旅游地情感波动与游记数量分布存在较大差异。12 月份是哈尔滨最受游客欢迎的月份，人流量最多，却不是游客积极情感最强烈的月份。存在游客积极情感强烈，但游客人数不多的月份，表现出这个月份旅游地的发展潜力。

由此总结得到网络游记情感的时间变化特征三为：网络游记情感随时间的波动与相应的游记数量并不相关，即传统的旅游地旺季和淡季的划分与游客的实际情感体验并不相关。

从上文分析可知，情感[好]、[乐]和[恶]在不同月份的波动程度较明显。因此，选取东北资源型的代表旅游地哈尔滨和南方海滨型的代表旅游地三亚，从[好]、[乐]和[恶]情感进行时间变化因素分析。情感[好]表达了游客对旅游地的尊敬、赞扬、相信、喜爱和祝愿，而[乐]表达了游客快乐、安心的情感，[恶]则表现了游客对旅游地的烦闷、憎恶、贬责、嫉妒和怀疑。由于情感[好]的均值远大于[乐]和[恶]，为更清晰地显示三种情感的变化趋势，将其分为两组绘制折线图，如图 6 和图 7 所示。

东北资源型的哈尔滨情感波动情况比南方海滨型的三亚更显著，说明不同类型旅游地的游客情感随时间变化，特征存在差异。东北资源型旅游地的游客在 1 月、7 月、11 月、12 月的积极情感[好]和[乐]较高，即夏冬季节，说明东北资源型旅游地具有夏季可避暑，冬天能玩冰雪的特点，能满足游客的预期。但同时冬季[恶]的情感很高，旅游地可以从冬季月份的游记中

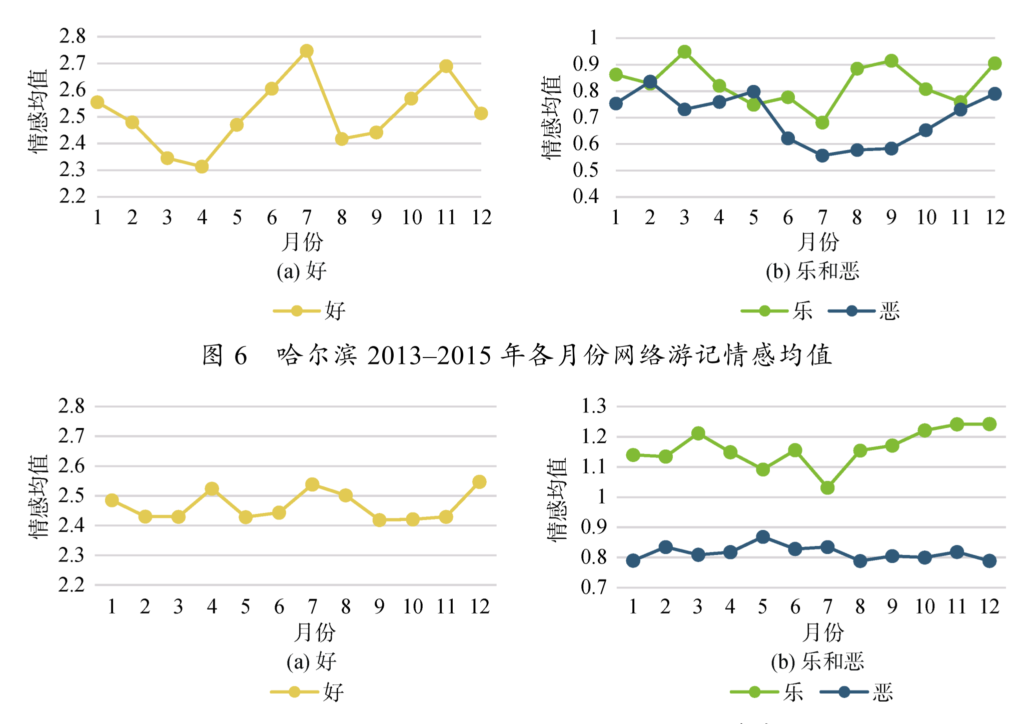


图 6　哈尔滨 2013–2015 年各月份网络游记情感均值

图 7　三亚 2013 年–2015 年各月份网络游记情感均值

挖掘游客产生[恶]情感的原因，提高旅游服务。而 4 月、5 月份情感[好]和[乐]都低于平均值，2 月、3 月[好]和[乐]情感一般，而[恶]情感上升，说明东北资源型旅游地缺乏对春季旅游产品开发，让游客失望而归。

与之相比，南方海滨型的三亚[好]的情感均值全年的平均值为 2.47，高于哈尔滨，且情感波动没有哈尔滨大，说明游客在三亚的体验普遍比哈尔滨好，旅游发展更稳定成熟。由图 7 可见，三亚 5 月–6 月[好]和[乐]情感都下降，[恶]情感明显上升。对于南方海滨型的旅游地，5 月–6 月天气炎热，而且常常伴随着台风下雨，对游客旅游体验的有很大影响。12 月[好]和[乐]情感都很高，对于喜欢嬉戏游玩的游客，推荐选择[乐]情感高的 10 月–12 月期间出游，而偏好休闲观光的游客，选择[好]情感高的 4 月、7 月、8 月、12 月期间旅游。

6　结 语

网络游记正在成为一种旅游方面重要的信息来源，一方面为潜在旅游者提供参考，另一方面为商家提供商机。本文从时间层面对旅游资源进行分析，提出面向网络游记时间特征的情感分析模型，该模型主要包括网络游记文本内容及旅游时间数据的采集、游记文本预处理、情感标注、按时间段统计游记情感特征分值、游记情感时间特征的结果分析 5 大模块。该模型从时间维度上分析游客对旅游地情感变化规律，基本规律如下:

(1) 在 7 类情感中，[好]的情感均值在各旅游地的各月份中总是最高，远高于其他情感，较为稳定。

(2) 在 7 类情感中，情感[好]、[乐]和[恶]在不同月份的波动程度较大，[怒]、[哀]、[惧]、[惊]的波动很小。

(3) 情感随时间的波动与相应游记数量并不相关，即传统的旅游地旺季和淡季的划分与游客的实际情感体验并不相关。

通过对 4 个不同类型旅游地的游记进行分析，发现符合以上规律。对于旅游地[好]和[乐]情感强度都高于平均值的月份，游客对旅程体验的满意度比较高，旅游资源具有很大的开发空间。对于[恶]较高的月份，不推荐给旅游者，建议旅游地管理者对该时段游客游记进行内容分析，了解游客厌恶的原因，并采取相关改善措施。对于游客积极情感强烈，但游客人数不多的月份，建议潜在游客可选择在该月份前往旅游。因此应用该模型能通过分析不同时间段游记的情感波动

情况，发现游客对旅游地情感倾向随时间变化的规律，进而为旅游管理者、潜在旅游者信息获取提供新的信息参考渠道。

本文不足之处在于通过游记的方式能获得丰富的信息，但其中可能会存在部分情感表达不充分的情况，如对旅游地很不满的游客很少会发表游记。在后续的研究中将考虑加入其他类型的信息来源进行分析，进一步完善旅游地情感时间特征的规律，使研究结论更具有参考价值。

参考文献：

[1] 杨立公，朱俭，汤世平. 文本情感分析综述[J]. 计算机应用，2013，33(6)：1574-1578. (Yang Ligong, Zhu Jian, Tang Shiping. Survey of Text Sentiment Analysis [J]. Journal of Computer Applications, 2013, 33(6): 1574-1578.)

[2] Hu M, Liu B. Mining and Summarizing Customer Reviews[C]//Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2004.

[3] 朱嫣岚，闵锦，周雅倩，等. 基于 HowNet 的词汇语义倾向计算[J]. 中文信息学报，2006，20(1)：14-20. (Zhu Yanlan, Min Jin, Zhou Yaqian, et al. Semantic Orientation Computing Based on HowNet[J]. Journal of Chinese Information Processing, 2006, 20(1): 14-20.)

[4] 史伟，王洪伟，何绍义. 基于微博平台的公众情感分析[J]. 情报学报，2012，31(11)：1171-1178. (Shi Wei, Wang Hongwei, He Shaoyi. Study on Public Sentiment Based on Microblogging Platform[J]. Journal of the China Society for Scientific and Technical Information, 2012, 31(11): 1171-1178.)

[5] Pang B, Lee L, Vaithyanathan S. Thumbs up?: Sentiment Classification Using Machine Learning Techniques[C]// Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. 2002.

[6] 徐军，丁宇新，王晓龙. 使用机器学习方法进行新闻的情感自动分类[J]. 中文信息学报，2007，21(6)：95-100. (Xu Jun, Ding Yuxin, Wang Xiaolong. Sentiment Classification for Chinese News Using Machine Learning Methods [J]. Journal of Chinese Information Processing, 2007, 21(6): 95-100.)

[7] Banić L, Mihanović A, Brakus M. Using Big Data and Sentiment Analysis in Product Evaluation[C]// Proceedings of the 36th International Convention on Information & Communication Technology Electronics & Microelectronics (MIPRO). 2013.

[8] 王素格，吴苏红. 基于依存关系的旅游景点评论的特征–观点对抽取[J]. 中文信息学报，2012，26(3)：116-121. (Wang Suge, Wu Suhong. Feature-Opinion Extraction in Scenic Spots Reviews Based on Dependency Relation [J]. Journal of Chinese Information Processing, 2012, 26(3): 116-121.)

[9] 郑文英. 旅行目的地中文评论的情感分析研究[D]. 哈尔滨：哈尔滨工业大学，2010. (Zheng Wenying. Sentiment Analysis of Travel Destination Reviews in Chinese[D]. Harbin: Harbin Institute of Technology, 2010.)

[10] 金程. 游客情感的动态性及其变化机制研究[D]. 广州：华南理工大学，2015. (Jin Cheng. Research on the Dynamics and Change Mechanism of Tourist Emotions [D]. Guangzhou: South China University of Technology, 2015.)

[11] 胡传东，李露苗，罗尚焜. 基于网络游记内容分析的风景道骑行体验研究——以 318 国道川藏线为例[J]. 旅游学刊，2015，30(11)：99-110. (Hu Chuandong, Li Lumiao, Luo Shangkun. Cycling Tourists' Experience of Scenic Byways Based on Content Analysis of Travel Blogs: A Case Study of the Sichuan-Tibet Section of National Highway 318[J]. Tourism Tribune, 2015, 30(11): 99-110.)

[12] 于静. 基于微博大数据的游客情感及时空变化研究[D]. 西安：陕西师范大学，2015. (Yu Jing. Research on Tourist Emotion and Spatio-temporal Variation Based on Microblog Big Data[D]. Xi'an: Shaanxi Normal University, 2015.)

[13] Li Q, Wu Y, Wang S, et al. VisTravel: Visualizing Tourism Network Opinion from the User Generated Content [J]. Journal of Visualization, 2016, 19(3): 1-14.

[14] 火车采集器[CP/OL]. [2016-11-15]. http://www.locoy.com/. (LocoySpider [CP/OL]. [2016-11-15]. http://www.locoy.com/.)

[15] 蚂蜂窝[DB/OL]. [2016-11-15]. http://www.mafengwo.cn/. (mafengwo.com [DB/OL]. [2016-11-15]. http://www.mafengwo.cn/.)

[16] 贺涛. 面向中文博客的信息采集与倾向性检索[D]. 合肥：中国科学技术大学，2009. (He Tao. Research on Chinese Blog Information Gathering and Opinion Retrieval [D]. Hefei: University of Science and Technology of China, 2009.)

[17] 杜振雷. 面向微博短文本的情感分析研究[D]. 北京：北京信息科技大学，2013. (Du Zhenlei. Sentiment Analysis Towards Microblog Short Text [D]. Beijing: Beijing Information Science and Technology University, 2013.)

[18] Standard Deviation [EB/OL]. [2016-11-15]. https://en.wikipedia.org/wiki/Standard_deviation.

作者贡献声明：

程翠琼，徐健：提出研究思路，设计研究方案；

程翠琼: 进行实验, 采集、清洗和分析数据;
程翠琼, 徐健: 论文起草及最终版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据见期刊网络版 http://www.infotech.ac.cn。

[1] 程翠琼. 旅游地网络游记情感分值.xlsx. 4 个旅游地网络游记的情感分值数据.
[2] 程翠琼. 旅游地情感分均值结果.xlsx. 4 个旅游地情感分均值结果数据.

收稿日期: 2016-10-07
收修改稿日期: 2016-11-30

A Sentiment Analysis Model Based on Temporal Characteristics of Travel Blogs

Cheng Cuiqiong　Xu Jian
(School of Information Management, Sun Yat-Sen University, Guangzhou 510006, China)

Abstract: [**Objective**] This study aims to find the temporal-distribution patterns of tourists' attitudes towards their destinations through sentiment analysis of travel blogs. [**Context**] More and more tourists collect information on their destinations from travel blogs, which provide enormous business opportunities. [**Methods**] We proposed a sentiment analysis model based on temporal characteristics of travel blogs. It includes the following modules: data collection, preprocessing, identifying sentiment words, weight calculation, and analysis. The model was examined with four types of travel blogs. [**Results**] The number of post with "good" emotion was always higher than others each month. The volatility of "good", "happiness" and "disgust" emotion was the highest in different months. The volatility emotion over time was not correlated to the number of related travel blogs. There is no relationship between the peak/off seasons and the emotion of tourists. [**Conclusions**] The proposed model could identify the changing of tourist sentiment over time, which provides new information for tourism managers and potential visitors.

Keywords: Travel Blogs　Sentiment Analysis　Sentiment Lexicon　Temporal Characteristics